



The ImageCLEF 2011 plant images classification task

Hervé Goëau, Pierre Bonnet, Alexis Joly, Nozha Boujemaa, Daniel Barthélémy, Jean-François Molino, Philippe Birnbaum, Elise Mouysset, Marie Picard

► To cite this version:

Hervé Goëau, Pierre Bonnet, Alexis Joly, Nozha Boujemaa, Daniel Barthélémy, et al.. The ImageCLEF 2011 plant images classification task. ImageCLEF 2011, Sep 2011, Amsterdam, Netherlands. hal-00642197

HAL Id: hal-00642197

<https://inria.hal.science/hal-00642197>

Submitted on 17 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The ImageCLEF 2011 plant images classification task

Hervé Goëau¹, Pierre Bonnet², Alexis Joly¹, Nozha Boujemaa¹, Daniel Barthelemy², Jean-Francois Molino², Philippe Birnbaum², Elise Mouysset³, and Marie Picard³

¹ INRIA, IMEDIA team, France, alexis.joly@inria.fr,
<http://www-rocq.inria.fr/imedia/>

² CIRAD, UMR AMAP, France, pierre.bonnet@cirad.fr,
<http://amap.cirad.fr/fr/index.php>

³ Telabotanica, France, elise@tela-botanica.org, <http://www.tela-botanica.org/>

Abstract. ImageCLEFs plant identification task provides a testbed for the system-oriented evaluation of tree species identification based on leaf images. The aim is to investigate image retrieval approaches in the context of crowdsourced images of leaves collected in a collaborative manner. This paper presents an overview of the resources and assessments of the plant identification task at ImageCLEF 2011, summarizes the retrieval approaches employed by the participating groups, and provides an analysis of the main evaluation results.

Keywords: ImageCLEF, plant, leaves, images, collection, identification, classification, evaluation, benchmark

1 Introduction

Convergence of multidisciplinary research is more and more considered as the *next big thing* to answer profound challenges of humanity related to health, biodiversity or sustainable energy. The integration of life sciences and computer sciences has a major role to play towards managing and analyzing cross-disciplinary scientific data at a global scale. More specifically, building accurate knowledge of the identity, geographic distribution and uses of plants is essential if agricultural development is to be successful and biodiversity is to be conserved. Unfortunately, such basic information is often only partially available for professional stakeholders, teachers, scientists and citizens, and often incomplete for ecosystems that possess the highest plant diversity. A noticeable consequence, expressed as the *taxonomic gap*, is that identifying plant species is usually impossible for the general public, and often a difficult task for professionals, such as farmers or wood exploiters and even for the botanists themselves. The only way to overcome this problem is to speed up the collection and integration of raw observation data, while simultaneously providing to potential users an easy and efficient access to this botanical knowledge. In this context, content-based visual identification of plant's images is considered as one of the most promising

solution to help bridging the taxonomic gap. Evaluating recent advances of the IR community on this challenging task is therefore an important issue.

This paper presents the *plant identification task* that was organized within ImageCLEF 2011⁴ for the system-oriented evaluation of visual based plant identification. This first year pilot task was more precisely focused on tree species identification based on leaf images. Leaves are far from being the only discriminant visual key between tree species but they have the advantage to be easily observable and the most studied organ in the computer vision community. The task was organized as a classification task over 70 tree species with visual content being the main available information. Additional information only included contextual meta-data (author, date, locality name) and some EXIF data. Three types of image content were considered: leaf *scans*, leaf photographs with a white uniform background (referred as *scan-like* pictures) and unconstrained leaf's *photographs* acquired on trees with natural background. The main originality of this data is that it was specifically built through a citizen sciences initiative conducted by TelaBotanica⁵, a French social network of amateur and expert botanists. This makes the task closer to the conditions of a real-world application: (i) leaves of the same species are coming from distinct trees living in distinct areas (ii) pictures and scans are taken by different users that might not used the same protocol to collect the leaves and/or acquire the images (iii) pictures and scans are taken at different periods in the year.

2 Task resources

2.1 The Pl@ntLeaves dataset

Building effective computer vision and machine learning techniques is not the only side of the taxonomic gap problem. Speeding-up the collection of raw observation data is clearly another crucial one. The most promising approach in that way is to build real-world collaborative systems allowing any user to enrich the global visual botanical knowledge [14]. To build the evaluation data of ImageCLEF plant identification task, we therefore set up a *citizen science* project around the identification of common woody species covering the Metropolitan French territory. This was done in collaboration with TelaBotanica social network and with researchers specialized in computational botany.

Technically, images and associated tags were collected through a crowd-sourcing web application [14] and were all validated by expert botanists. Several cycles of such collaborative data collection and taxonomical validation occurred. Scans of leaves were first collected over two seasons, between July and September 2009 and between June and September 2010 thanks to the work of active contributors from TelaBotanica social network. The idea of collecting only scans during this first period was to initialize the training data with limited noisy background and to focus on plant variability rather than mixed plant and view conditions

⁴ <http://www.imageclef.org/2011>

⁵ <http://www.tela-botanica.org/>

Genus Species	Vernacular name	Genus Species	Vernacular name
Acer campestre L.	Field Maple	Ligustrum vulgare L.	European Privet
Acer monspessulanum L.	Montpellier Maple	Magnolia grandiflora L.	Southern magnolia
Acer negundo L.	Box elder Maple	Malus sylvestris Mill.	European Wild Apple
Acer opalus Mill.	italian Maple	Melia azedarach L.	Persian Lilac
Acer platanoides L.	Norway Maple	Nerium oleander L.	Nerium oleander
Acer pseudoplatanus L.	Sycamore Maple	Olea europaea L.	Olive
Aesculus hippocastanum L.	Horse-chestnut	Paliurus spina-christi Mill.	Jerusalem Thorn
Ailanthus altissima (Mill.) Swingle	tree of heaven	Phillyrea angustifolia L.	Mock Privet
Alnus glutinosa (L.) Gaertn.	Black Alder	Pistacia lentiscus L.	Mastic
Arbutus unedo L.	Strawberry Tree	Pistacia terebinthus L.	Turpentine tree
Betula pendula Roth	Silver Birch	Pittosporum tobira (Thunb.) W.T.Aiton	Japanese mock-orange
Broussonetia papyrifera (L.) Vent.	Paper Mulberry	Platanus x hispanica Mill. ex Münchh.	Hybrid Plane
Carpinus betulus L.	European hornbeam	Prunus mahaleb L.	St Lucie cherry
Castanea sativa Mill.	Sweet Chestnut	Prunus serotina Ehrh.	Wild Black Cherry
Celtis australis L.	European nettle tree	Prunus spinosa L.	Blackthorn
Cercis siliquastrum L.	Judas Tree	Punica granatum L.	Pomegranate
Cornus mas L.	European Cornel	Quercus coccifera L.	Kermes Oak
Corylus avellana L.	Common Hazel	Quercus ilex L.	Holm Oak
Cotinus coggygria Scop.	Eurasian smoketree	Quercus petraea Liebl.	Durmast Oak
Crataegus azarolus L.	Azarole	Quercus pubescens Willd.	Pubescent Oak
Crataegus monogyna Jacq.	Hawthorn	Rhamnus alaternus L.	Italian Buckthorn
Diospyros kaki L.f.	Kaki Persimmon	Rhamnus cathartica L.	Purgive Buckthorn
Eriobotrya japonica (Thunb.) Lindl.	Japanese medlar	Rhus typhina L.	Staghorn Sumac
Fagus sylvatica L.	European Beech	Robinia pseudoacacia L.	Black Locust
Ficus carica L.	Common fig	Salix caprea L.	Goat Willow
Fraxinus angustifolia Vahl	Narrow-leaved Ash	Sambucus nigra L.	Black Elder
Fraxinus excelsior L.	European Ash	Sophora japonica L.	Pagoda Tree
Fraxinus ornus L.	Manna Ash	Sorbus aria (L.) Crantz	Common Whitebeam
Ginkgo biloba L.	Ginkgo	Sorbus domestica L.	Service Tree
Gleditsia triacanthos L.	Honey locust	Sorbus torminalis (L.) Crantz	Wild Service Tree
Ilex aquifolium L.	European Holly	Syringa vulgaris L.	Common Lilac
Juglans nigra L.	Eastern Black walnut	Tilia cordata Mill.	Small-leaved Lime
Juglans regia L.	Common walnut	Viburnum lantana L.	Wayfaring Tree
Laburnum anagyroides Medik.	Common Laburnum	Viburnum tinus L.	Laurestine
Laurus nobilis L.	Laurel Tree	Vitis agnus-castus L.	Vitex

Fig. 1. List of tree species included in the Pl@ntLeaves dataset

variability. This allowed to collect 2228 scans over 55 species. A public version of the application ⁶ was then opened in October 2010 and additional data were collected up to March 2011. The new collected images were either *scans*, or photographs with uniform background (referred as *scan-like photos*), or unconstrained *photographs* with natural background. They involved 15 new species from the previous set of 55 species. The Pl@ntLeaves dataset used within ImageCLEF finally contained 5436 images with 3070 scans, 897 scan-like photos and 1469 photographs. Figure 2 displays samples of these 3 image types for 4 distinct tree species. The full list of species is provided in Figure 1.

2.2 Pl@ntLeaves metadata

Each image of Pl@ntLeaves dataset is associated with the following meta-data:

- *Date* upload date of the image
- *Type* (acquisition type: *scan*, *scan-like* or *photograph*)

⁶ http://combraille.cirad.fr:8080/demo_plantscan/












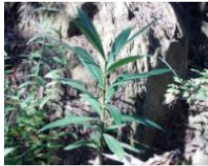
Genus Species	Scans	Scan-like photos	Photographs
<i>Acer Negundo</i> L.			
<i>Celtis Australis</i> L.			
<i>Cercis siliquastrum</i> L.			
<i>Nerium Oleander</i> L.			

Fig. 2. Illustration of the 3 image type categories for 4 species

- *Content* content type: single leaf, single dead leaf or foliage (several leaves on tree visible in the picture)
- *Taxon* full taxon name (sub-regnum, regnum, class, division, order, family, genus, species)
- *VernacularNames* French or English vernacular names
- *Author* name of the author of the picture
- *Organization* name of the organization of the author
- *Locality* locality name (a district or a country division or a region)
- *GPSLocality* GPS coordinates of the observation

These meta-data are stored in independent xml files, one for each image. Figure 3 displays an example image with its associated xml data.

Additional but partial meta-data information can be found in the image's EXIF, and might include the camera or the scanner model, the image resolution and

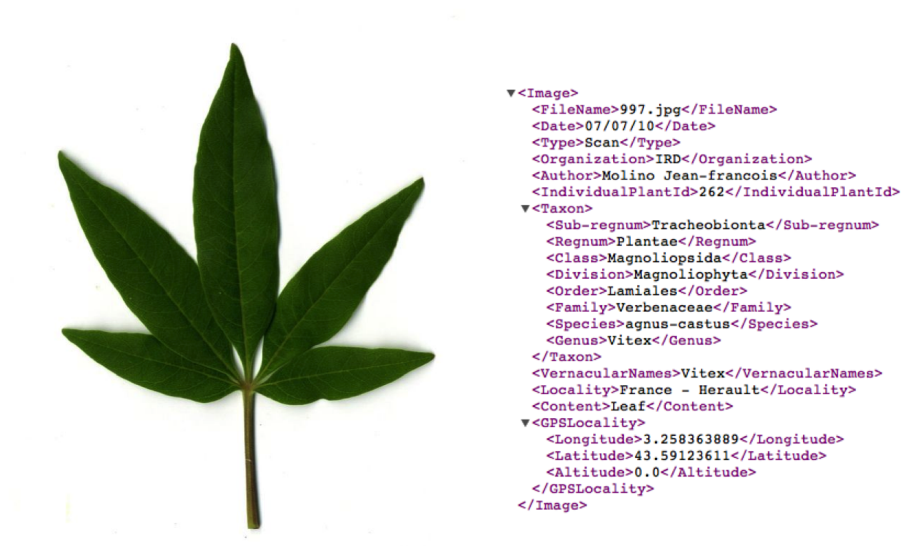


Fig. 3. An image of Pl@ntLeaves dataset and its associated metadata

dimension, the optical parameters, the white balance, the light measures, etc.

2.3 Pl@ntLeaves variability

The main originality of Pl@ntLeaves compared to previous leaf datasets, such as the Swedish dataset [16] or the Smithsonian one [9], is that it was built in a collaborative manner through a citizen sciences initiative. This makes it closer to the conditions of a real-world application: (i) leaves of the same species are coming from distinct trees living in distinct areas (ii) pictures and scans are taken by different users that might not used the same protocol to collect the leaves and/or acquire the images (iii) pictures and scans are taken at different periods in the year. Intra-species visual variability and view conditions variability are therefore more stressed-out which makes the identification more realistic but more complex. Figures 4 to 9 provide illustrations of the intra-species visual variability over several criteria including leaf's color, leaf's global shape, leaf's margin appearance, number and relative positions of leaflets and number of lobes. On the other side, Figure 10 illustrates the light reflection and shadows variations of *scan-like photos*. It shows that this acquisition protocol is actually very different than pure scans. Both share the property of a limited noisy background but scan-like photos are much more complex due to the lighting conditions variability (flash, sunny weather, etc.) and the *unflatness* of leaves. Finally, the variability of unconstraint *photographs* acquired on the tree and with



Fig. 4. Color variation of *Cotinus coggygia* Scop. (Eurasian smoketree)



Fig. 5. Global shape variation of *Corylus avellana* L. (European Hazel)

natural background is definitely a much more challenging issue as illustrated in Figure 11.

3 Task description

The task was evaluated as a supervised classification problem with tree species used as class labels.

3.1 Training and Test data

A part of Pl@ntLeaves dataset was provided as training data whereas the remaining part was used later as test data. The training subset was built by randomly selecting 2/3 of the **individual plants** of each species (and not by randomly splitting the images themselves). So that pictures of leaves belonging to the same individual tree cannot be split across training and test data. This prevents identifying the species of a given tree thanks to its own leaves and that makes the task more realistic. In a real world application, it is indeed much unlikely that a user tries to identify a tree that is already present in the training data. Detailed statistics of the composition of the training and test data are provided in Table 1.



Fig. 6. Leaf's margin variation of *Quercus ilex* L. (Holm oak)



Fig. 7. Number of leaflets variation of *Fraxinus angustifolia* Vahl (Narrow-leaved Ash)

		Nb of pictures	Nb of individual plants	Nb of contributors
Scan	Train	2349	151	17
	Test	721	55	13
Scan-like	Train	717	51	2
	Test	180	13	1
Photograph	Train	930	72	2
	Test	539	33	3
All	Train	3996	269	17
	Test	1440	99	14

Table 1. Statistics of the composition of the training and test data

3.2 Task objective and evaluation metric

The goal of the task was to associate the correct tree species to each test image. Each participant was allowed to submit up to 3 runs built from different methods. As many species as possible can be associated to each test image, sorted by decreasing confidence score. Only the most confident species was however used in the primary evaluation metric described below. But providing an extended ranked list of species was encouraged in order to derive complementary statistics (e.g. recognition rate at other taxonomic levels, suggestion rate on top k species,



Fig. 8. Leaflets relative position variation of *Vitis agnus-castus* L. (Judas Tree)



Fig. 9. Number of lobes variation of *Ficus carica* L. (Common Fig)

etc.).

The primary metric used to evaluate the submitted runs was a *normalized classification rate* evaluated on the 1st species returned for each test image. Each test image is attributed with a score of 1 if the 1st returned species is correct and 0 if it is wrong. An average *normalized* score is then computed on all test images. A simple mean on all test images would indeed introduce some bias with regard to a real world identification system. Indeed, we remind that the Pl@ntLeaves dataset was built in a collaborative manner. So that few contributors might have provided much more pictures than many other contributors who provided few. Since we want to evaluate the ability of a system to provide correct answers to all users, we rather measure the mean of the average classification rate per author. Furthermore, some authors sometimes provided many pictures of the same individual plant (to enrich training data with less efforts). Since we want to evaluate the ability of a system to provide the correct answer based on a single plant observation, we also decided to average the classification rate on each individual plant. Finally, our primary metric was defined as the following average classification score S :

$$S = \frac{1}{U} \sum_{u=1}^U \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{n=1}^{N_{u,p}} s_{u,p,n} \quad (1)$$



Fig. 10. Light reflection and shadows variation of scan-like photos of *Magnolia Grandiflora* (Southern Magnolia)



Fig. 11. Variability of unconstrained photographs of *Acer platanoides* (Norway Maple)

U : number of users (who have at least one image in the test data)

P_u : number of individual plants observed by the u -th user

$N_{u,p}$: number of pictures taken from the p -th plant observed by the u -th user

$s_{u,p,n}$: classification score (1 or 0) for the n -th picture taken from the p -th plant observed by the u -th user

It is important to notice that while making the task more realistic, the normalized classification score also makes it more difficult. Indeed, it works as if a bias was introduced between the statistics of the training data and the one of the test data. It highlights the fact that bias-robust machine learning and computer vision methods should be preferred to train such real-world collaborative data. Finally, to isolate and evaluate the impact of the image acquisition type (scan, scan-like, photograph), a normalized classification score S was computed for each type separately. Participants were therefore allowed to train distinct classifiers, use different training subsets or use distinct methods for each data type.

4 Participants and techniques

A total of 8 groups submitted 20 runs, which is a successful participation rate for a first year pilot task on a new topic. Participants were mainly academics, specialized in computer vision and multimedia information retrieval, coming from all around the world: Australia (1), Brazil (1), France (2), Romania (1), Spain

(1), Turkey (1) and UK (1). We list below the 8 participants and give a brief overview of the techniques they used to run the plant identification task. We remind here that ImageCLEF benchmark is a system-oriented evaluation and not a formal evaluation of the underlying methods. Readers interested by the scientific and technical details of any of these methods should refer to the CLEF2011 working note of each participant.

IFSC (3 runs) [1] The two best runs obtained by this participant (IFSC_USP_run2 & IFSC_USP_run1) are mainly based on a new shape boundary analysis method they introduced recently [11]. It is based on the complex network theory [10]. A shape is modeled into a small-world complex network and it uses degree and joint degree measurements in a dynamic evolution network to compose a set of shape descriptors. This method is claimed to be robust, noise tolerant, scale invariant and rotation invariant and proved to provide better performances than Fourier shape descriptors, curvature-based descriptors, Zernike moments and multiscale fractal dimensions.

LIRIS (4 runs) [2] This participant also used a classification scheme based on shape boundary analysis. The main originality however is that they used a model-driven approach for the segmentation and shape estimation. Their four runs differ in the parameters of the method.

UAIC (3 runs) [3] This participant was the only one trying to benefit from metadata associated to the images (location, date, author, etc.). They submitted therefore 3 runs to evaluate the contribution of metadata compared to using visual content only. Their first run (UAIC2011_Run01) is based on visual content only, the second one (UAIC2011_Run02) uses only metadata based features in the classification process, and the third one uses both (UAIC2011_Run03).

SABANCI (1 run) [4] This participant submitted only one run based on a supervised classification approach using support vector machine (SVM) and a combination of 3 visual features: a new texture feature they introduced in ??, a shape boundary feature based on FFT coefficients [] and some basic global color features. This is the only method combining leaf boundary information with other texture and color features.

INRIA (2 runs) [5] This participant submitted two runs based on two radically different methods. Their second run (inria_imedia_plantnet_run2) is based on a shape boundary feature, called DFH [17], that they introduced in 2006. Their first run (inria_imedia_plantnet_run1) is more surprising for a supervised classification task of leaves since it is based on local features matching with rigid geometrical models. Such generalist method is usually more dedicated to large-scale retrieval of rigid objects and this is the only participant who used such approach.

RMIT (2 runs) [6] RMIT mainly focused on comparing two distinct machine learning algorithms: nearest-neighbor classifier (RMIT_run1) decision tree (RMIT_run2).

DAEDALUS (1 run) [7] This participant used a generalist image retrieval framework based on SIFT features and a nearest-neighbor classifier.

KMIMMIS (4 runs) [8] This participant also used a generalist image retrieval framework based on local features and a nearest-neighbor classifier. They compared different configurations: basic clustered SIFT with 1-NN label transfer (kmimmis_run1 and kmimmis_run4), simple edge and corner point detection with 1-NN vote label transfer (kmimmis_run2), simple edge and corner point detection with 10-NN vote label transfer (kmimmis_run3).

Over all runs, the most frequently used class of methods is *shape boundary analysis*. 8 runs among 20 are based on some boundary shape features. This is not surprising since state-of-the-art methods addressing leaf-based identification in the literature are mostly based on leaf segmentation and shape boundary features [12, 15, 13, 17, 11]. On the other side, it was a good news that the majority of the runs were based on other various approaches so that more relevant conclusions can be expected.

5 Results

5.1 Global analysis

Figures 12, 13 and 14 present the normalized classification scores of the 20 submitted runs for each of the three image types. Alternatively, Figure 15 presents the overall performances averaged over the 3 image types. Table 2 finally presents the same results but with detailed numerical values.

A first global remark is that, as expected, the performances are degrading with the complexity of the acquisition image type. Scans are more easy to identify than scan-like photos and unconstrained photos are much more difficult. This is can be easily seen in Figure 15 where the relative scores of each image type are highlighted by distinct colors.

A second global remark is that no method provide the best score for all image types. None of the run even belongs to the top-3 runs of all image types (as shown in Table 2). This is somehow disappointing from the genericity point of view but not surprising regarding the nature of the different image types. One could expect that scans and scan-like photos lead to similar conclusions but this is actually not the case. The only runs that give quite stable and good performances over the three image types are the two runs of IFSC based on complex network shape boundary analysis method (IFSC_USP_run1 & IFSC_USP_run2). This justifies their excellent ranking when averaging the classification scores over

the three image types. All other methods fail in providing as good results for the unconstrained photographs as shown in Figure 14. This score gap between IFSC run and others has however to be mitigated by a bias introduced by the author normalization of the classification score. Indeed, their high score is mostly due to excellent performances on the images of one of the 3 contributors. All these images are very similar and less cluttered than the average of the unconstrained photos (actually they are all close-up of a juda's tree leaf). But still, it seems that this is the only method that have at least perform very well on these images.

A third important remark is that shape boundary analysis methods do not provide the best results on the scan images whereas they are usually considered as being state-of-the-art on such data. They all provide good classification scores between 48% and 56% but they are consistently outperformed by two more generic image retrieval approaches (as shown in Figure 12). The best score is achieved by INRIA's run using large-scale matching of local features with rigid geometrical models (`inria_imedia_plantnet_run1`, 68% classification rate). This suggests that modeling leaves as part-based rigid and textured objects might be an interesting alternative to shape boundary approaches that do not characterize well margin details, ribs or limb texture. Second best score on scans is obtained by the run of SABANCI (`Sabanci-run1`) which uses a supervised classification approach based on support vector machine (SVM) and a combination of 3 global visual features. This suggests that combining shape boundary features with other color and shape textures is also a promising direction. Global conclusions on the scan-like photos are quite different (Figure 13). The best score is obtained by INRIA's run (`inria_imedia_plantnet_run2`) purely based on a global shape boundary feature (DFH [17]). It is followed closely by the four runs of LIRIS also based on boundary shape features but using a model-driven approach for the segmentation and the shape estimation. Then come the two runs of INRIA and SABANCI that ranked first on the scan images (the first one based on rigid objects matching and the second one training combined features) and finally the shape boundary method of IFSC. One conclusion is that shape boundary methods appear to provide more stable results for scans and scan-like photos. On the other side the rigid object matching method of INRIA degrades much more from scans to scan-like pictures. This can be explained by the fact that it is more discriminant regarding the leave's morphology but less robust to light reflections and shadows. These lighting variations might also explain the degrading performances of the combined features used by SABANCI.

5.2 About using metadata

Using metadata to help the identification, and more particularly geo-tags, is definitely something that has to be studied. We were therefore very enthusiastic to see the runs of UAIC, aimed at evaluating the potential contributions of `Pl@ntLeaves` metadata. Unfortunately, results clearly show that adding metadata degrades their identification performances. Metadata alone even give a null

Run id	Participant	Scans	Scan-like	Photographs	Mean
IFSC_USP_run2	IFSC	0,562	0,402	0,523	0,496
inria_imedia_plantnet_run1	INRIA	0,685	0,464	0,197	0,449
IFSC_USP_run1	IFSC	0,411	0,430	0,503	0,448
LIRIS_run3	LIRIS	0,546	0,513	0,251	0,437
LIRIS_run1	LIRIS	0,539	0,543	0,208	0,430
Sabancı-run1	SABANCI	0,682	0,476	0,053	0,404
LIRIS_run2	LIRIS	0,530	0,508	0,169	0,403
LIRIS_run4	LIRIS	0,537	0,538	0,121	0,399
inria_imedia_plantnet_run2	INRIA	0,477	0,554	0,090	0,374
IFSC_USP_run3	IFSC	0,356	0,187	0,116	0,220
kmimmis_run4	KMIMMIS	0,384	0,066	0,101	0,184
kmimmis_run1	KMIMMIS	0,384	0,066	0,040	0,163
UAIC2011_Run01	UAIC	0,199	0,059	0,209	0,156
kmimmis_run3	KMIMMIS	0,284	0,011	0,060	0,118
UAIC2011_Run03	UAIC	0,092	0,163	0,046	0,100
kmimmis_run2	KMIMMIS	0,098	0,028	0,102	0,076
RMIT_run1	RMIT	0,071	0,000	0,098	0,056
RMIT_run2	RMIT	0,061	0,032	0,043	0,045
daedalus_run1	DAEDALUS	0,043	0,025	0,055	0,041
UAIC2011_Run02	UAIC	0,000	0,000	0,042	0,014

Table 2. Normalized classification scores for each run and each image type. Top 3 results per image type are highlighted in bold

classification success rate. Besides technical details that could probably slightly improve such species filtering based on metadata, it still shows that Pl@ntLeaves metadata might be intrinsically not very useful for identification purposes. The main reason is probably that the geographic spread of the data is limited (French mediterranean area). So that most species of the dataset might be identically and uniformly distributed in the covered area. Geo-tags would be for sure more useful at a global scale (continent, countries). But at a local scale, the geographical distribution of plants is much more complex. It usually depends on localized environmental factors such as sun exposition or water proximity that would require much more data to be modeled.

5.3 Performances per species

To evaluate which species are more difficult to identify than others, we averaged the performances over the runs of all participants (for each specie). It is however difficult to understand precisely the score variations. They can be due to morphological variations but also to different view conditions or other statistical bias in the data such as the number of training images. Figure 16 presents the obtained graph for the scan images only (in order to limit view conditions bias). The only global trend we discovered so far is that single leaves are on average

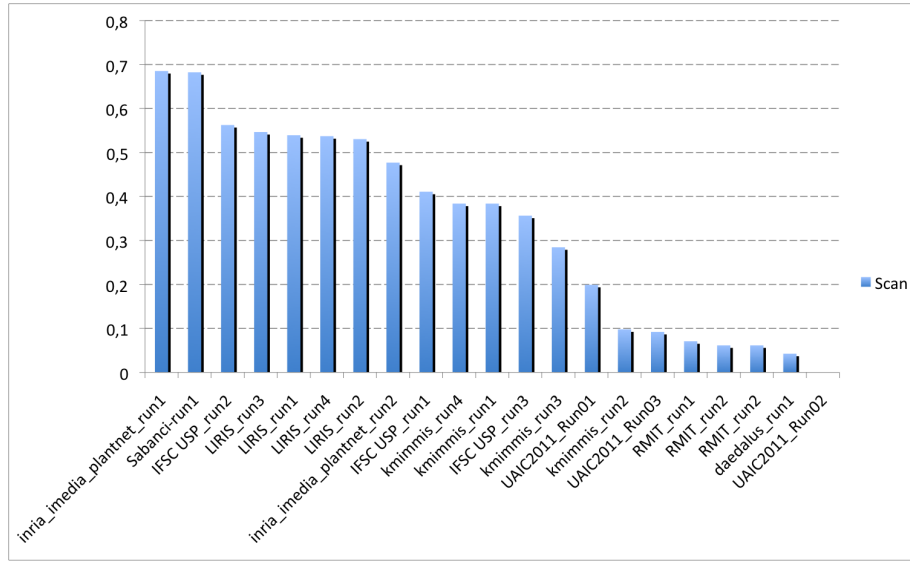


Fig. 12. Normalized classification scores for scan images

easier to identify than compound leaves.

5.4 Performances per image

To qualitatively assess which kind of images causes good results and which one makes all methods failed, we sorted all test pictures by the number of runs in which they were correctly identified. The obtained ranking confirms that scan images are much easier for the identification. 99% of 100-top ranked images are actually scans (with 11 to 17 successful runs). Figure 17 displays the 4 best identified images (with 17/20 successful runs). They all are very standard leaf images similar to the one found in books illustrations. On the other side, 260 images were not identified by any run with a majority of unconstrained photos (63 scans, 27 scan-like photos, 168 unconstrained photographs). The scans and scan-like photos belonging to this category of *most difficult* images are very interesting. As illustrated in Figure 18, most of them correspond to outlier leaves with defaults or unusual morphology (ill or torn leaves, missing leaflets, etc.). Figure 18 displays 8 of them.

6 Conclusions

This paper presented the overview and the results of ImageCLEF 2011 plant identification testbed. A challenging collaborative dataset of tree's leaf images

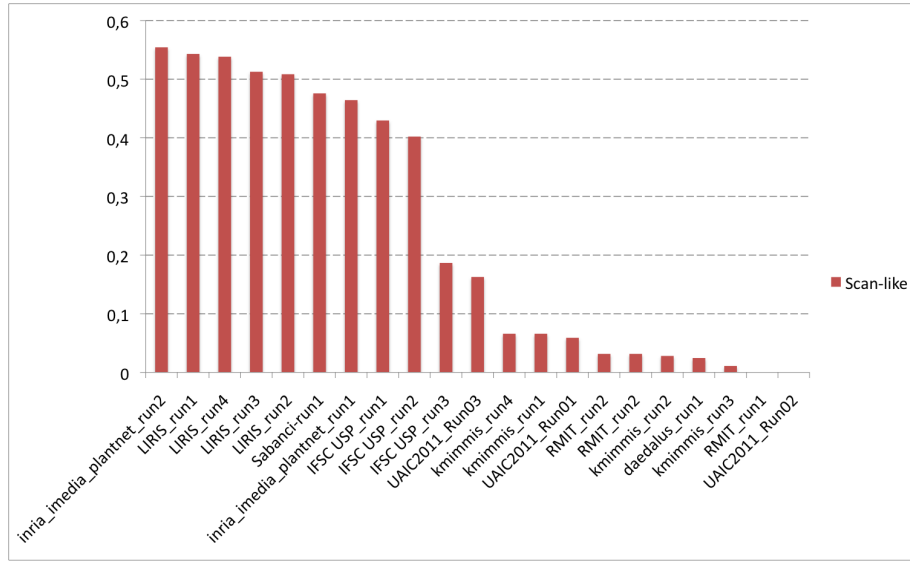


Fig. 13. Normalized classification scores for scan-like photos

was specifically built for this evaluation and 8 participants coming from different countries submitted a total of 20 runs. A first conclusion was that identification performances are close from mature when using scans or photos with uniform background but that unconstrained photos are still much more challenging. More data and evaluation are clearly required to progress on such data. Another important conclusion was that state-of-the-art methods based on shape boundary analysis were not the best ones on leaf scans. Better performances were notably obtained with a local features matching technique usually more dedicated to the large-scale retrieval of rigid objects. On the other side, shape boundary analysis methods remain better on scan-like photos due to their better robustness to light reflections and shadow effects. This suggests that combining shape boundary features with part-based rigid object models might be an interesting direction. Adding texture and color information also showed some improvements. On the contrary, using additional metadata such as geo-tags was not concluding on the evaluated dataset. Probably because the geographic spread of the data was limited.

Acknowledgement

This work was funded by the Agropolis fundation through the project Pl@ntNet (<http://www.plantnet-project.org/>) and the EU through the CHORUS+ Coordination action (<http://avmediasearch.eu/>)

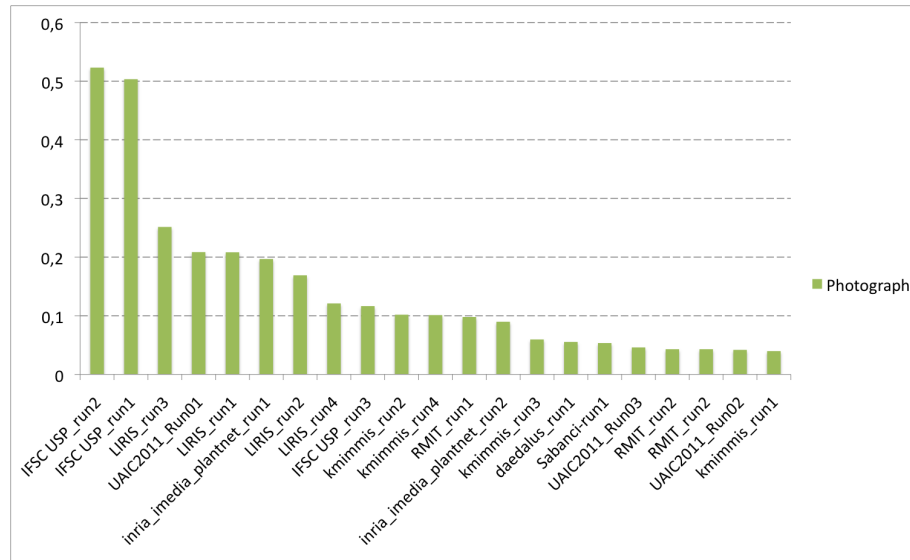


Fig. 14. Normalized classification scores for photographs

References

1. In: Working notes of CLEF 2011 conference (2011)
2. In: Working notes of CLEF 2011 conference (2011)
3. In: Working notes of CLEF 2011 conference (2011)
4. In: Working notes of CLEF 2011 conference (2011)
5. In: Working notes of CLEF 2011 conference (2011)
6. In: Working notes of CLEF 2011 conference (2011)
7. In: Working notes of CLEF 2011 conference (2011)
8. In: Working notes of CLEF 2011 conference (2011)
9. Agarwal, G., Belhumeur, P., Feiner, S., Jacobs, D., Kress, J.W., R. Ramamoorthi, N.B., Dixit, N., Ling, H., Mahajan, D., Russell, R., Shirdhonkar, S., Sunkavalli, K., White, S.: First steps toward an electronic field guide for plants. *Taxon* 55, 597–610 (2006)
10. Albert, R.Z.: Statistical mechanics of complex networks. Ph.D. thesis, Notre Dame, IN, USA (2001)
11. Backes, A.R., Casanova, D., Bruno, O.M.: A complex network-based approach for boundary shape analysis. *Pattern Recognition* 42(1), 54 – 67 (2009)
12. Belhumeur, P., Chen, D., Feiner, S., Jacobs, D., Kress, W., Ling, H., Lopez, I., Ramamoorthi, R., Sheorey, S., White, S., Zhang, L.: Searching the worlds herbaria: A system for visual identification of plant species. In: *ECCV*, pp. 116–129 (2008)
13. Bruno, O.M., de Oliveira Plotze, R., Falvo, M., de Castro, M.: Fractal dimension applied to plant identification. *Information Sciences* 178(12), 2722 – 2733 (2008)
14. Goeau, H., Joly, A., Selmi, S., Bonnet, P., Mouysset, E., Joyeux, L.: Visual-based plant species identification from crowdsourced data. In: *Proceedings of ACM Multimedia 2011* (2011)

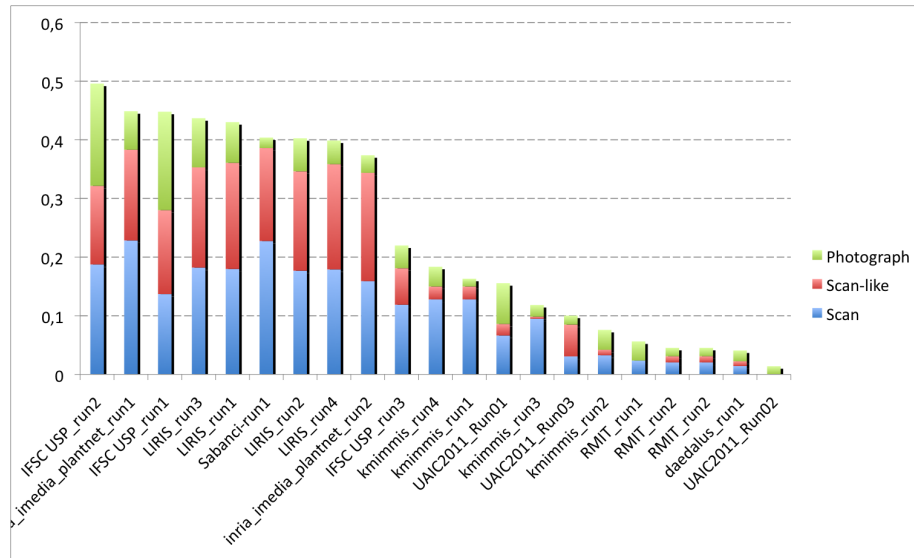


Fig. 15. Normalized classification scores averaged over all image types

15. Neto, J.C., Meyer, G.E., Jones, D.D., Samal, A.K.: Plant species identification using elliptic fourier leaf shape analysis. *Computers and Electronics in Agriculture* 50(2), 121 – 134 (2006)
16. Söderkvist, O.J.O.: Computer Vision Classification of Leaves from Swedish Trees. Master's thesis, Linköping University, SE-581 83 Linköping, Sweden (September 2001), liTH-ISKY-EX-3132
17. Yahiaoui, I., Herv, N., Boujemaa, N.: Shape-based image retrieval in botanical collections. In: *Advances in Multimedia Information Processing - PCM 2006*, vol. 4261, pp. 357–364 (2006)

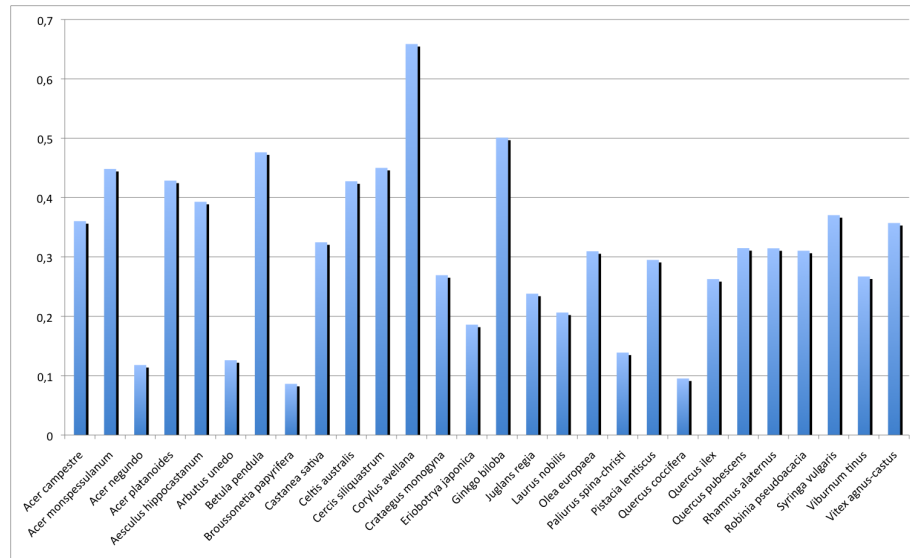


Fig. 16. Mean classification score per species averaged over all participant runs - scan test images only

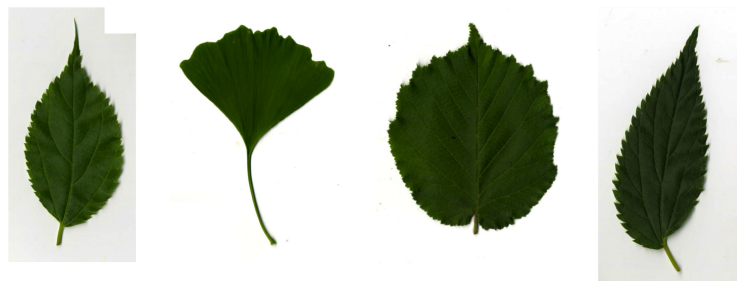


Fig. 17. 4 best identified images (17/20 successful runs)



Fig. 18. **Top:** 4 of the most difficult test scans (0/20 successful runs) - **Bottom:** 4 of the most difficult scan-like test photos (0/20 successful runs)